DATA SCIENCE AND ANALYTICS
PROFESSIONAL BOOTCAMP

INTRODUCTORY COURSE

Introduction to Data Science and Analytics

MODULE COMPANION

MODULE OUTLINE

Lesson 01: Program Overview

Become familiar with the program, learning outcomes, hands-on projects, and program resources.

Lesson 02: Getting Started with Data Science and Analytics

1. Explore the wide variety of roles that are possible.
2. Identify characteristics of data science and analytics jobs.
3. Examine specific companies that use analytics, machine learning, and business intelligence.
4. List ways that different industries such as insurance, education, and banking are using analytics and data science.

Lesson 03: Tools and Software

1. Identify the programming languages used in data science, analytics, machine learning, etc.
2. Identify the different types of software that data professionals use.
3. Define GitHub and explain how it's used by data professionals.
4. Build a data analytics portfolio in GitHub.

1.1.1 PROGRAM OVERVIEW

Become familiar with the program, learning outcomes, hands-on projects, and program resources.

Discussion

What attracted you to data science and analytics and made you decide to pursue this program and a career in this field?

What Are the Lesson Goals?

Become familiar with the program, learning outcomes, hands-on projects, and program resources.

Why Are They Important?

- Understand how you play an active role to complete this learning experience successfully.
- Utilize program resources.

Program Overview

The skills required in the professional landscape are evolving. Technological advancements are moving quickly, and, quite simply, there are not nearly enough people equipped with the skills needed to fill the open data professional positions.

This market-driven program addresses this massive deficit in the workforce. It addresses the skills gap in the workforce. You will get to experience scenarios and assignments that closely align with data challenges currently faced by businesses.

How are we going to get there?

Program Path

Introductory Course
SQL and Databases
Statistics and Probability
Data Storytelling
Milestone Project 1: Building and Presenting Data Stories
Python Programming
Data Wrangling
Visual Communication
Advanced SQL Programming
Milestone Project 2: Data Integration, Preparation, Reporting, and Presentation
Business Intelligence
Big Data
Machine Learning
Applied AI
Milestone Project 3: Capstone Project: Delivering Insights and Presentations

Program Goals

During this bootcamp, you will learn how to:

- Organize business requirements into data requirements; see the big picture; and work on a variety of data ingestion, cleaning, and analysis assignments.
- Work responsibly with data.
- Leverage the power of artificial intelligence.
- Build visualizations that are accessible to everyone to help the business achieve its goals.
- Work on a variety of data ingestion, cleaning, and analysis assignments.

Discussion

What do you wish to accomplish by attending the Data Science and Analytics Bootcamp?

Review and Wrap-Up

In this lesson, you learned about:

- Various careers in data
- Different types of job roles
- What type of work data professionals do

1.1.2 GETTING STARTED WITH DATA SCIENCE AND ANALYTICS

- **What is data science and analytics?**
- **Careers in data science and analytics**
- **Characteristics of data science and analytics careers**

What Are the Lesson Goals?

1. Demonstrate an understanding of what data science and analytics is.
2. Describe business solutions in different industries offered by data science and analytics.
3. Recognize the difference between various careers in the data profession.
4. Describe the daily job tasks of a data analyst.

The main objective of this lesson is to show a wide range of examples of data science and analytics solutions across different industries. The application of data science/analytics takes place in nearly every industry and sector. More and more companies are using data to their advantage.

Analytics help us extract useful information from large amounts of data to make better, data-based decisions.

Why Are They Important?

It's important to gain an understanding and appreciation of the possibilities of data science and analytics. Throughout this lesson, you will see there are a variety of industries and businesses that leverage their data to meet their business needs. Even within a business, each department can utilize analytics in different ways to serve the department and company at large.

What Is Data?

Data is information in digital form that is stored on a computer or server and used as a basis for analysis reasoning, discussion, or calculation.

It is information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful.

What Is Data Science

“Data science combines the scientific method, math and statistics, specialized programming, advanced analytics, AI, and even storytelling to uncover and explain the business insights buried in data.”—IBM

“Data science is a multidisciplinary approach to extracting actionable insights from the large and ever-increasing volumes of data collected and created by today’s organizations. Data science encompasses preparing data for analysis and processing, performing advanced data analysis, and presenting the results to reveal patterns and enable stakeholders to draw informed conclusions.”—IBM

From Data to Insight

Data insights refer to the strategy businesses use to collect, analyze, and act on data related to the business and its clients and then make better decisions. A data scientist’s role is to extract insights from data using many strategies, including predictive analytics, artificial intelligence (AI), machine learning, and deep learning models. Let’s establish an understanding of these key terms before we dig deeper.

Predictive Analytics

Analytics help us extract useful information from large amounts of data to make better data-based decisions. Many companies use analytics where data is stored in central repositories and analysts query this data and build reports and dashboards. Companies use analytics to help them gain new insights and make data-driven decisions.

“Predictive analytics is a branch of advanced analytics that makes predictions about future outcomes using historical data combined with statistical modeling, data mining techniques and machine learning. Companies employ predictive analytics to find patterns in this data to identify risks and opportunities.”—IBM

Artificial Intelligence

“In computer science, the term artificial intelligence (AI) refers to any human-like intelligence exhibited by a computer, robot, or other machine. In popular usage, artificial intelligence refers to the ability of a computer or machine to mimic the capabilities of the human mind—learning from examples and experience, recognizing objects, understanding and responding to language, making decisions, solving problems—and combining these and other capabilities to perform functions a human might perform, such as greeting a hotel guest or driving a car.”—WorldNoor

Machine Learning

The number of companies utilizing data science, analytics, and machine learning is growing. Most companies are using data science and machine learning methods to transform data into useful insights.

Netflix can take your viewing habits and transform them into recommendations. The more information you tell Netflix about what you like to watch, the more Netflix can recommend things you like. This feature is based on “recommendation engines.”

Spotify uses techniques called natural language processing (NLP) and collaborative filtering (a type of recommender).

Deep Learning

“Deep learning is a subset of machine learning in which multi-layered neural networks—modeled to work like the human brain—'learn' from large amounts of data. Within each layer of the neural network, deep learning algorithms perform calculations and make predictions repeatedly, progressively 'learning' and gradually improving the accuracy of the outcome over time.”—Tech Grid

Data Science in Business

Regardless of industry, it’s helpful to think about how data scientists and analytics can be used in various functional areas within a business. Each department can leverage data to meet their needs in a different way, such as:

- Human resources analytics (people analytics)
- Marketing analytics
- Operations analytics
- Production and manufacturing
- Supply chain analytics
- Customer service analytics

Insurance and Banking

Insurance and banking companies use analytics for fraud detection, risk analysis, claims and processing analysis, and process optimization. Fraud detection techniques are designed to prevent money being obtained illegally, such as via money laundering, forging checks, and using stolen credit cards. Mastercard has incorporated machine learning and AI to detect fraud based on the size, location, and time of the transaction. Predictive analysis can analyze both numeric and text-based data to find questionable or suspicious activity.

Health Care

Health care is another industry with a significant amount of data. The data is complex, from medical and pharmacy claims to patient registries and data within electronic health record (EHR) systems. This data can be used to:

- Analyze patient cases
- Improve patient health and well-being
- Identify and predict disease
- Identify at-risk patients

Law Enforcement

Law enforcement has evolved significantly to use heatmaps and predictive analytics to target and fight crime. While most of what you see on TV is fiction, detective screens with multiple graphical solutions and spreadsheets are accurate.

DataRobot

DataRobot is an artificial intelligence (AI) platform that helps automate machine learning. It can:

- Analyze, model, and present data
- Customize reports using Python
- Democratize data

In addition to DataRobot, another company called HMH takes raw assessment data, state assessment data, grades, and other qualitative data to analyze and provide summarized reports about student achievement and performance. HMH works with K–12 school districts across the U.S., Canada, and Mexico to provide textbooks and software to help students learn. They share their insights with school district leaders.

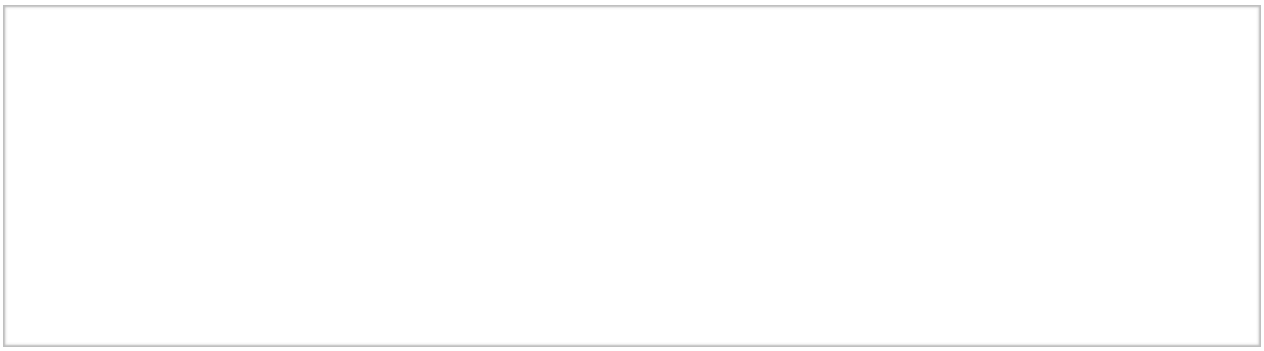
Education

K–12, higher education, and continuing education programs are making data-driven decisions regarding student achievement, student outcomes, and identifying and responding to at-risk students. This is helpful for identifying students who are struggling. Adaptive learning platforms use machine learning and AI techniques to deliver content tailored to each individual student based on their needs and achievements.

Activity: Discussion

What are some examples of companies using data science and analytics? Provide examples of professional experiences. Share stories of how your organization might use analytics and data science. It's possible you might not know whether organizations are using data science

and analytics. Think about how companies use their data, which can help uncover the ways that companies are using these techniques.



Poll

Which of the following are types of solutions that can be found in banking and insurance?

A. Identifying risks

B. Identifying fraudulent activity

C. Processing claims

D. All the above

Poll Feedback

Which of the following are types of solutions that can be found in banking and insurance?

Identifying risks and fraudulent activity and processing claims are all types of solutions that can be found in the banking and insurance industries.

Breakout: Examples of Data Science and Analytics Solutions

In this activity, you'll explore the wide variety of ways data is being transformed with analytics and data science. In fact, many businesses have been able to transform their data into useful information for better decision-making, better customer experience and customer service, and increased efficiency.

Pick an industry you are interested in and search for analytics solutions within it. Answer the following questions:

- How is data science and analytics used in this industry?

- What are the types of problems the industry is trying to solve using data science and analytics?

DSA Career Options

There are many possible careers in which you can work with data. In this program, we are focused primarily on data analyst and data scientist roles, but we want you to recognize there are a significant number of roles that work with data in their daily jobs. You may have seen these different job titles, but they all work with data.

During the course, we'll discuss various aspects of data analytics, business intelligence, data science, and machine learning. There are numerous roles across data, whether you're working as a data analyst, BI developer, or data scientist. All these roles work with data as part of their roles.

It's important to remember there are many different types of jobs available for people skilled in working with data. Some may gravitate to working on data visualization, business intelligence, reporting, and so forth, while others may want to dive into topics like applied AI and machine learning. Technical product managers (TPMs) often work with other professionals in a team setting to help manage projects and keep people on target to complete their work and tasks. We also have scrum masters who help facilitate daily stand-up meetings and other agile rituals (such as sprint retros, sprint planning, and so on).

The key takeaway is that many job roles need to understand data and may work with other more technical staff but will still need to understand data enough to facilitate and communicate with other team members.

Role Comparison

The work of a data professional varies greatly. Depending on the role, you may be required to work with internal and external customers to build visualizations. On the other hand, you might work with tools like Python to develop ETL pipelines or investigate and analyze data to build prediction models. At some point, you might find yourself gathering, organizing, and curating data. Other examples of daily work include:

- Working with customers to understand their needs and data requirements
- Giving presentations based on data analyses
- Managing data-related projects
- Writing code in SQL, Python, R, and other languages
- Working in teams and individually

Data Scientist vs. Data Analyst

Data Scientist Role:
Data science encompasses preparing data for analysis and processing, performing advanced data analysis, and presenting the results to reveal patterns and enable stakeholders to draw informed conclusions.

Data Analyst Role:

- Analyze company or industry data to discover new insights.
- Build reports to share with decision-makers.
- Work with data from multiple sources.

There are various data analyst roles that focus on different challenges depending on the specific departments or functions within a business. Analyst roles include:

- Marketing
- Human resources (HR)
- Operations

For example, marketing analysts use their expertise in digital marketing to report on website traffic to ensure companies don't waste money on campaigns that aren't bringing visitors to their sites.

Breakout Rooms

Pair up with a classmate and discuss your interests in data science.

- What types of data-related roles interest you?
- What characteristics do you share with a data scientist or analyst?
- Conduct a web search for the types of roles that interest you and find a job

description. You can use websites like ***Dice.com***, ***Indeed.com***, and ***LinkedIn.com***. Detail the job here.

Activity: Career Services

Complete the following:

- Career Services Prep Module
- Career Services Survey
- Job Exploration Worksheet
- Submit your resume and LinkedIn URL.

Review and Wrap-Up

- Various industries and companies that use analytics
- Areas within businesses that use analytics

1.1.3 TOOLS AND SOFTWARE

- **Data software**
- **Programming languages**
- **Data visualization software**
- **Building a data analytics portfolio in GitHub**

Review

Previously, you were exposed to the cool ways that different industries are using data science and analytics to solve interesting problems. Today, we're going to look closely at the tools and software that data professionals use to solve these problems and mimic what you will use on the job.

What Are the Lesson Goals?

1. Identify the programming languages used in data science.
2. Identify the different types of software that data professionals use.
3. Define GitHub and explain how it's used by data professionals.
4. Build a data analytics portfolio in GitHub.

Why Are They Important?

As a data professional, you will use a variety of tools to perform analysis on large data sets, as well as visualize and present the insights that solve these interesting problems businesses face. In this lesson, you'll get an overview of what tools are available and see what you can expect to learn later in the program.

Data Software

Data professionals use various types of software. There is software for data visualization, ensuring data quality, preparing and wrangling data, building data warehouses and catalogs, and maintaining databases. Throughout this program, you will learn how to use different software to perform these same tasks on data, so you are prepared to perform a variety of tasks on the job. Pictured is a JavaScript library for manipulating documents based on data.

Microsoft Excel

Spreadsheets like Microsoft Excel and Google Sheets help data professionals analyze, clean, and visualize data. Spreadsheets are also useful in creating a table (also called a matrix).

Languages

Programming languages allow us to program our computers. Later in the program, you will use SQL and Python. Numerous languages are used because no two jobs are exactly alike. There's been a long-standing debate about whether Python or R is better for data science. They are both very useful, and each language has its own strengths and weaknesses, which you will dive deeper into in the Python and R courses. Other programming languages used by data professionals include:

- C++
- Scala
- Javascript

SQL

Structured Query Language (SQL) is one of the most common languages used by data professionals. Databases used in tandem with SQL are Microsoft SQL Server, MySQL, Oracle, and Hadoop.

Python

Python is an open-source programming language that is high level and works as a general-purpose language; it is most often compared to Ruby, JavaScript, and Scheme.

Quiz

What is the most popular language used to query databases?

A. Python
B. Scala
C. R
D. SQL

Quiz Feedback

What is the most popular language used to query databases?

D: SQL or Structured Query Language is a language used to query different types of relational databases. It is used with databases such as MySQL, PostgreSQL, Microsoft SQL Server, Oracle, and many others.

Tableau

Tableau is one example of data visualization software that takes raw data and transforms it into a visual form. You can create a variety of charts and graphs (line, bar, heatmap, etc.) to create dashboards. Other data visualization software include:

- Microsoft PowerBI
- Qlik
- Domo

GitHub

GitHub is a place to store and showcase all your data projects as a portfolio for potential employers.

"GitHub is a website and cloud-based service that helps developers store and manage their code, as well as track and control changes to their code."

How Does It Work?

Module 1.8 will further elaborate on GitHub.

- Users send files and folders (with files) to the GitHub website.
- GitHub supports collaborating on a project, so users might add/retrieve files to/from GitHub constantly as new versions are made.
- The GitHub website is a cloud server. The cloud is a place people can store files through online access.
- You can store any file format in GitHub (text documents, images, videos, scripts with code, etc.).
- The folders may be nested inside other folders that are inside other ones.
- The free version of GitHub has the privacy level set to public. Anyone can see your GitHub portfolio. If you pay, you can have a private GitHub portfolio. Public is the norm, as most users have a public GitHub portfolio.

Repos

A repository is a place to store your projects.

Examples:

- Data-Analyst-Portfolio by mkumar7
- data_science_portfolio by melvfnz
- Data Science Portfolio by Arch Desai

Activity 1.1.3: Build Your GitHub Repo!

Build your data analyst portfolio on GitHub to store and showcase all your data projects and use it as a portfolio for potential employers.

Steps

- Create a GitHub account.
- Create a new repository.
- Customize the project summary page.
- Populate your portfolio with data projects you work on throughout the Introductory Course.

Download the Activity Guide

Download the Activity Guide from Canvas and follow the instructions on how to complete this project. Refer to Activity 1.1.3.

Review and Wrap-Up

- What data science and analytics is
- Careers in data science and analytics
- Tools and software that data professionals use
- How to build a portfolio in GitHub

Next Steps

- Assigned activities
- Data analyst portfolio: GitHub tasks
- Other reminders